

Topological classifier for detecting the emergence of epileptic seizures

Piangerelli Marco, Matteo Rucco and Emanuela Merelli *

November 16, 2016

Abstract

In this work we study how to apply topological data analysis to create a method suitable to classify EEGs of patients affected by epilepsy. The topological space constructed from the collection of EEGs signals is analyzed by Persistent Entropy acting as a global topological feature for discriminating between healthy and epileptic signals. The Physionet data-set has been used for testing the classifier.

Keywords: Complex systems, Brain, Epilepsy, Topological Data Analysis, Persistent Entropy, Classifier, Time series

1 Introduction

According to the World Health Organization (WHO), epilepsy is a chronic brain disorder characterized by recurrent seizures, which may vary from a brief lapse of attention or muscle jerks to severe and prolonged convulsions. The seizures are caused by sudden, usually brief, excessive electrical discharges in a group of brain cells (neurons)¹. Epilepsy can be classified according to the portion of the involved neural cortex: the *focal* epilepsy seizures are spatially confined either in small part of a lobe or in the whole hemisphere, while the *generalized* epilepsy seizures spread in the whole brain. In both cases epileptic seizures are a spontaneous hyper-synchronous activity of clusters of neurons [16]. It is worth mentioning that the human brain is a complex self adaptive system, composed of billion of non-identical neurons, entangled in loops of non-linear interactions, those determining the brain behaviors; the epilepsy is an example [33]. In our perspective, identifying the onset of a neural hyper-synchronization is likely to discover patterns of information expressed

*Computer Science Division, School of Science and Technology, University of Camerino, Camerino, Italy

¹www.who.int/topics/epilepsy/

by a network of interactions in the space of neurons. Unfortunately, to the best of our knowledge, monitoring the electrical activity of each single neuron is not feasible, but it is possible to capture the electrical activity of the whole brain (or of a part of it). Currently, the standard technique used to record the signals through the positioning of certain number of electrodes on the scalp, is the electroencephalogram (EEG), whose direct observation helps neurologists in diagnosing epilepsy; the use of methods for the automatic diagnosis is still far from being a reality. In the last decades several methods have been proposed in literature – linear and non-linear analysis [35, 17], applications of chaos theory [13] and dynamical systems modelling [29, 12], but none of them is suitable. The intrinsic non-linearity and the non-stationarity of EEG signals request for more suitable methods capable of extracting global information, both structural and functional, characterizing the brain cluster of neurons involved in the hyper-synchronous activity.

To this end, starting from the analysis of the EEG signals we must reconstruct the abstract structural component that represents the interactions among neuronal clusters involved in the brain activity and from the analysis of the evolution of the structural component (the abstract model of the brain) based on windows of observations, we must be able to identify the emergence of an hyper-synchronous activity of the brain neurons. In a short communication [18], the authors have proposed the use of topological data analysis (TDA for short) for EEG signals, to characterize the structural component and the emergence of epileptic seizures. This approach based on TDA has been previously introduced within the TOPDRIM² project for the analysis of fMRI signals [23]. In this paper we describe a method suitable to automatically classify EEG signals that record epileptic seizures from those recording the activity of an healthy brain. This step is very important and preliminary to reconstruct the abstract structural and functional components of the brain because it allows us to characterize the transition phase from an healthy to an epileptic state of the system. The proposed method uses the topological data analysis (TDA) combining two powerful instruments, persistent homology and persistent entropy measure for analysing the set of EEG signals and construct a topological classifier. TDA consists of a set of algorithms for investigating the higher dimensional relations hidden in data sets through the construction of simplicial complex; it finds its theoretical justification in the branch of mathematics called topology. The key-concept in TDA is the concept of *persistent homology*: a procedure for counting, through a process called *filtration*, the higher dimensional persistent holes. TDA has been applied in various fields among which computer vision and sensor coverage problem [7], signal analysis [22, 27], biological problems [5, 14, 23] and in studying the spreading of contagions in social and biological systems [32]. Persistent Entropy is a Shannon-like entropy for the measurement of the information discovered during the filtration process within the TDA. Shannon entropy [25] is the well known concept in thermodynamics as well as in information theory. The contribution of this work is twofold: from the one hand, we describe how the Persistent Entropy (PE for short) can be used

²Topology driven methods for complex systems: www.topdrim.eu

to discriminate the epileptic state versus non epileptic states. From the other hand, we conjecture that the Vietoris-Rips filtration helps to understand which region plays the role of trigger for an epileptic seizure. The paper is organized as it follows: in Section 2 we introduce the new paradigm of TDA and in Section 3 the TDA-based methodology used for the implementation of the new classifier is explained; finally in Section 4 the results of the work are presented and discussed.

2 TDA: a new paradigm for data analysis

In this section we introduce the TDA paradigm as the tool for analysing the EEG data set and discovering elements that are correlated in a given a data-set by n -relations. Suppose we have a set of points G , i.e. data, embedded in a d -dimensional space \mathbb{D}^d and assume that those data were sampled by an unknown k -dimensional space \mathbb{V}^k with $k \leq d$, our task is to reconstruct the space \mathbb{V}^k from the data set G , as it is in solving the inverse problem.

For instance, traditional analysis techniques, such as manifold learning and principal components analysis (PCA) [21], or unsupervised clustering techniques [8], make some fundamental assumptions over the space \mathbb{V}^k : it is a manifold; it is locally Euclidean; its metric is well defined; it lacks of curvature but it is smooth [36]. Since real data-sets are not a vector space they violate those assumptions, as consequence we claim that \mathbb{V} and \mathbb{D} have to be reshaped to *topological spaces*. Intuitively a topological space is a set of elements that are equipped with a notion of *proximity* parameter that gives rise to a coordinate-free metric. Instead of using the points as the vertices of a combinatorial graph to capture local information, in TDA the data set is converted into a topological space to discover the higher dimensional hidden information through the *filtration* of a sequence of approximated topological spaces called simplicial complexes. Building a filtration can be seen as wearing a lens for examining the data set; different lenses, i.e. different filtrations, let extract different information from the topological space. In this paper we deal with two kinds of filtrations: the *Lower star filtration* and the *Vietoris - Rips filtration*

For the sake of completeness we introduce some formal concepts: A *topology* on a set X is a subset $T \subseteq 2^X$ such that:

- If $S_1, S_2 \in T$, then $S_1 \cap S_2 \in T$ (equivalent to: If $S_1, S_2, \dots, S_n \in T$ then $\cap_{i=1}^n S_i \in T$)
- If $\{S_j | j \in J\} \subseteq T$, then $\cup_{j \in J} S_j \in T$.
- $\emptyset, X \in T$.

The pair (X, T) of a set X and a topology T is a *topological space*. A *simplicial complex* is a kind of topological space constructed by the union of n -dimensional simple pieces in such a way that the common intersection of two pieces are lower-dimensional pieces of the same kind. More concretely, an abstract *simplicial complex* K is composed by a set K_0 of 0-simplices (also called vertices V , that can be thought as points in \mathbb{R}^n) and, for each

$k \geq 1$, a set K_k of k -simplices $\sigma = \{v_0, v_1, \dots, v_k\}$, where $v_i \in V$ for all $i \in \{0, \dots, k\}$, satisfying that:

- each k -simplex has $k + 1$ faces obtained removing one of its vertices;
- if a simplex σ is in K , then all faces of σ must be in K .

The underlying topological space of K is the union of the geometric realization of its simplices: points for 0-simplices, line segments for 1-simplices, filled triangles for 2-simplices, filled tetrahedra for 3-simplices and their n -dimensional counterparts for n -simplices (see Figure 1). We only consider finite (abstract) simplicial complexes with finite dimension, i.e., there exists an integer n (called the dimension of K) such that for $k > n$, $K_k = \emptyset$ and for $0 \leq k \leq n$, K_k is a finite set.

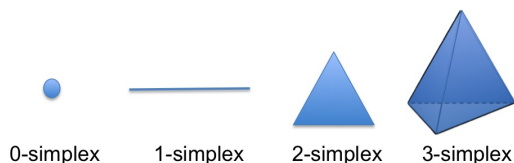


Figure 1: Main simplices.

A *filtration* on a simplicial complex K is a collection of subcomplexes $\{K(t) | t \in \mathbb{R}\}$ of K such that $K(t) \subset K(t')$ whenever $t \leq t'$. The filtration value of a simplex $\sigma \in K$ is the smallest t such that $\sigma \in K(t)$. A *filtered simplicial complex* is a simplicial complex equipped with a filtration.

As the filtration is the lens for reading the data it is worth understanding what kind of information exactly we are looking for. The answer is: the topological invariants or if you prefer the k -dimensional holes in the data. During the filtration process k -dimensional holes appear and disappear and some of them persist for a certain filtration-time: the short-lived are considered noise while the long-lived are the invariants we are looking for, see § 2.2. See [11] and [20] for an introduction to algebraic topology.

2.1 How to build simplicial complexes from data

As we mention above, to choose how to build a filtration is very important because it reflects how data points are converted into the simplicial complexes and thus converting set points into a topological space. Literature is plenty of techniques for obtaining simplicial complexes from data [9, 4, 3] and each of them depends on the particular application we are interested in. In this paper data are represented by multivariate time series (see Figure2). They are EEG signals of the *PhysioNet* database and freely accessible via web³. For performing an EEG, electrodes are pasted at some key points on the patients head

³<http://www.physionet.org>

following some schemes. The 10-20 system is the internationally recognized method to apply the location of electrodes in EEG recording. The 10-20 refers to the fact that actual distances between electrodes are either 10% or 20% of front-back or right-left distance of the skull (see Figure 3). The electrodes are assumed to record the electrical activity of the underlying cortical area that, in turn, is given by the interaction of the activity of all the neurons of that area. The EEG signals used in this study were collected at the Children’s Hospital Boston, and they consists of EEG recordings from pediatric subjects with intractable seizures. Subjects were monitored for up to several days following withdrawal of anti-seizure medication in order to characterize their seizures and assess their candidacy for surgical intervention. From the Physionet database we obtained 33 EEGs with at least one epileptic event and 33 records of EEG without epileptic events, by looking for patients with the same number of sensors and with recordings of the same length. Each record contains 23 one-dimensional signals with a sampling frequency of 256 Hz and of 921600 time points. Upon these data we build to kind of different complexes: the Piecewise

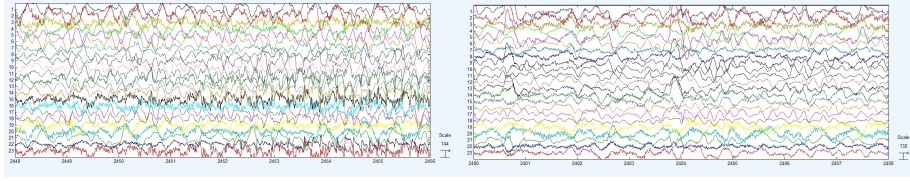


Figure 2: Example of an electric signal recorded by a patient with epileptic seizures, on the left, and from a healthy patient, on the right.

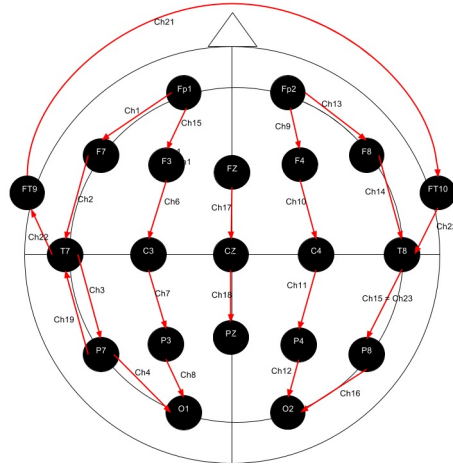


Figure 3: Graphical schema representing the positions of the sensors. The arrows correspond to the 23 potential differences that are recorded during the EEG.

complexes and the Vietoris-Rips complexes.

Piecewise complexes

Piecewise linear function (PL) is a powerful mathematical tool largely used for approximating signals. The task of measuring the similarity among piecewise linear functions (PLs) is still an open issue and a solution is strongly required in machine learning methods. A piecewise linear function is a real-valued function defined on the real numbers (see Figure 4).

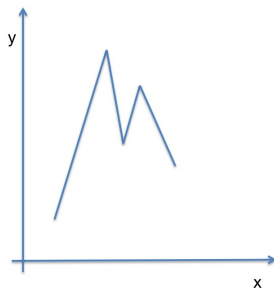


Figure 4: Graphical representation of a piecewise linear function. The graph is composed of straight-line sections.

In order to apply topological method to these functions, they must be equipped with a topology. In [28], the authors reported a new methodology that allows to associate a filtered simplicial complex to a PL. Then, Persistent Entropy, that is a topological summary is computed from the new simplicial complexes and it is used for the classification tasks of the topological spaces. Moreover, the authors have proven the stability theorem for the persistent entropy that is the formal method for comparing the persistent entropies of different simplicial complexes. Here we report on the theoretical methodology and one possible implementation. Given an unknown continuous signal $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}$, suppose that our input is the value of \tilde{f} on a finite set of points $S \subset \mathbb{R}^n$.

- Let K be a simplicial complex with real values specified at all vertices in S . E.g., if $S \subset \mathbb{R}$, then K is a line subdivided in segments with endpoints in D .
- Using linear extension over the cells of K , we obtain a piecewise linear (PL) function $f : K \rightarrow \mathbb{R}$ (being $f(u) = \tilde{f}(u)$ for $u \in S$). It is convenient to assume that f is *generic* by which we mean that the vertices have distinct function values. To ensure unique values, f may need to be perturbed. One way of doing this is to add a linear ramp to f (see [34, page 1650]). We can then order the vertices by increasing function value as $f(u_1) < f(u_2) < \dots < f(u_n)$.

- Now, the *lower star* of u_i can be computed which is the subset of simplices for which u_i is the vertex with maximum function value,

$$St_{-}u_i = \{\sigma \in St u_i : x \in \sigma \Rightarrow f(x) \leq f(u_i)\}.$$

The considered filtration is the *lower star filtration* of f (see [10, page 135]): $\emptyset = K_0 \subset K_1 \subset \dots \subset K_n = K$, in which K_i is the union of the first i lower stars.

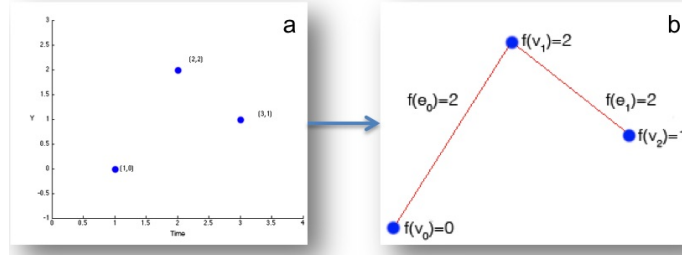


Figure 5: Graphical representation of the methodology that transforms a PL into a filtered simplicial complex. From left to right: a) The input signal formed by three time points, respectively with coordinates: $(1,0)$, $(2,2)$, $(3,1)$. b) The filtered simplicial complex formed by three 0-simplices: $\{v_0, v_1, v_2\}$ with filter values $f(v_0) = 0, f(v_2) = 1, f(v_1) = 2$ and two 1-simplices: $\{e_1, e_2\}$, with filter values $f(e_1) = f(e_2) = 2$, so the filter-value set is $F = \{0, 1, 2\}$.

For practical purposes, the methodology explained above can be translated in the following algorithm designed for analyzing a 2-dimensional plot (see Figure 5). Suppose that the first coordinate of a point in \mathbb{R}^2 represents time. Given a signal $S \subset \mathbb{R}^2$:

- order the points in S respect to their first coordinate (i.e., order the points in S by time);
- transform S into a filtered simplicial complex:
 - each point of S is a 0-simplex with filter equal to its second coordinate.
 - Each pair formed by two consecutive points in S : (x_i, y_i) and $(x_{i+1}, y_{i+1}) \in S$ where $x_i < x_{i+1}$, forms a 1-simplex σ with filter value $f(\sigma) = \max\{y_i, y_{i+1}\}$. Note that the resulting filtration is obtained by presenting at the beginning the simplices formed with the lowest second coordinate (i.e., the filter-value set F is obtained by spanning the Y -axis in a upward direction).

The resulting filtration is a *lower start filtration*.

Vietoris-Rips complexes

Vietoris-Rips filtration is a versatile tool in topological data analysis and it is used for studying point cloud data (PCD). More formally, Vietoris-Rips it is a sequence of simplices built on a metric space to add topological structure to an otherwise disconnected set of points. It is widely used because it encodes useful information about the topology of the underlying metric space. Two classical examples of abstract simplicial complexes are Čech complexes and Vietoris-Rips complexes (see [10, Chapter III]). Let V be a finite set of points in \mathbb{R}^n . The *Čech complex* of V and r denoted by $\check{C}_r(V)$ is the abstract simplicial complex whose simplices are formed as follows. For each subset S of points in V , form a closed ball of radius $r/2$ around each point in S , and include S as a simplex of $\check{C}_r(V)$ if there is a common point contained in all of the balls in S . This structure satisfies the definition of abstract simplicial complex. The *Vietoris-Rips complex* denoted as $VR_r(V)$ is essentially the same as the Čech complex. Instead of checking if there is a common point contained in the intersection of the $(r/2)$ -ball around v for all v in S , we may just check pairs adding S as a simplex of $\check{C}_r(V)$ if all the balls have pairwise intersections. We have $\check{C}_r(V) \subseteq VR_r(V) \subseteq \check{C}_{\sqrt{2}r}(V)$. See Fig.6.

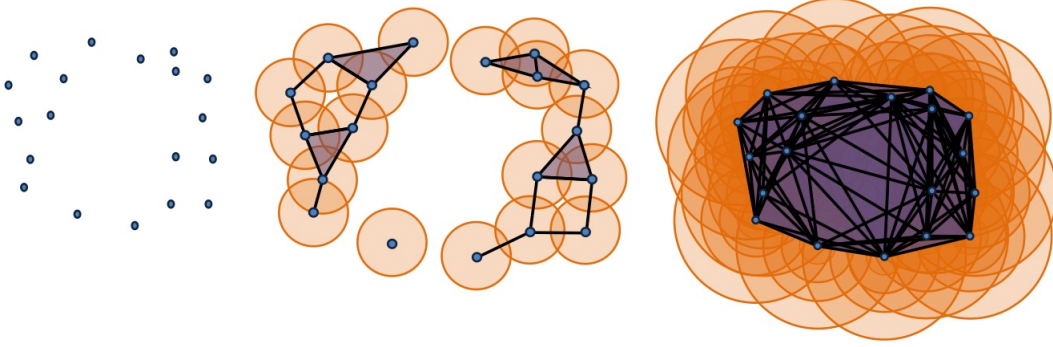


Figure 6: How to obtain Vietoris-Rips complexes from a PCD: starting from a PCD in a metric space, we surround each point with a sphere. All the spheres grow up simultaneously and equally. Each radius gives rise to new intersections. A new intersection of dimension k is equal to a $k-1$ simplices.

2.2 How to analyze simplicial complexes: persistent homology

A topological space is described by its own topological invariants. In case of the simplicial complex K it might be studied by homology. *Homology* is an algebraic machinery that counts the number of “ n -dimensional” holes in a topological space. Then, K is described by the dimension of its homological groups. These invariants are called Betti numbers. The value $Betti_k$, where $k \in \mathbb{N}$, is equal to the rank of the k -th homology group of

K . In particular, $Betti_0$ is the number of connected components, $Betti_1$ represents 2-dimensional holes, $Betti_2$ the voids in a 3-dimensional space, etc. . . . For instance, a k -dimensional empty sphere has all Betti numbers equal to zero except for $Betti_0 = Betti_k = 1$. Persistent homology is the combinatorial counterpart of homology. It takes as input a filtered simplicial complex, it does not mind how the filtered simplicial complex has been obtained. Persistent homology describes how the homology of K changes along filtration. A k -dimensional Betti interval, with endpoints $[t_{start}, t_{end})$, corresponds to a k -dimensional hole that appears at filtration time t_{start} and remains until time t_{end} . We refer to the holes that are still present at $t = t_{max}$ as *persistent topological features*, otherwise they are considered *topological noise* [1]. The set of intervals representing birth and death times of homology classes is called the *persistence barcode* associated to the corresponding filtration. Instead of bars, we sometimes draw points in the plane such that a point $(x, y) \in \mathbb{R}^2$ (with $x < y$) corresponds to a bar $[x, y)$ in the barcode. This set of points is called *persistence diagram* [37]. In Figure 7 we represent the same simplicial complexes of Figure 5 but also

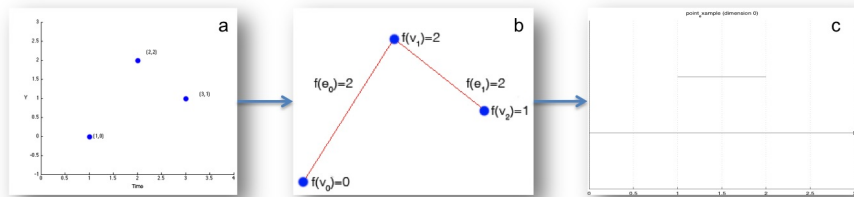


Figure 7: Graphical representation of the methodology that transforms a PL into a filtered simplicial complex. From left to right: a) The input signal. b) The corresponding filtered simplicial complex. c) The persistent barcodes.

equipped with its own persistent barcodes (box “c”). The persistent barcodes: at $F = 0$, there is only one topological feature corresponding to v_0 ; at $F = 1$, v_0 is still in the space but also a new component is introduced and it corresponds to v_2 ; eventually for $F = 2$, a new 0-simplex is added to the topological space (v_1) within the two 1-simplices e_1, e_2 where $e_1 = \{v_0, v_2\}$ and $e_2 = \{v_2, v_1\}$. From this filter value and successive, the space is described by only one persistent connected component, i.e. $\beta_0 = 1$. Visually there is only *one infinite line in the barcode*.

Persistent homology does not limit to count the number of higher dimensional holes but it also returns the simplices that are involved in the holes. These simplices are called *generators*. Even if there is not a rigorous justification, there are a bunch of experimental observations advocating that these generators plays a crucial role for describing the data under analysis. For example, in [19], the authors used TDA for studying the human immune systems and they recognized that the generators of persistent holes are exactly the antibodies involved in the immune memory. In [15], the authors used the Vietoris-Rips

complexes for analysing gene expression matrices and they discovered that the generators of the persistent holes are exactly the genes that are the responsible for certain pathologies.

2.3 Statistics for persistent homology: persistent entropy

In order to measure how much is ordered the construction of a filtered simplicial complex, a new entropy measure, the so-called *the persistent entropy*, has been defined in [26]. A precursor of this definition was given in [6] to measure how different bars of the barcode are in length. Here we recall the definition.

Persistent entropy. Given a filtered simplicial complex $\{K(t) : t \in F\}$, and the corresponding persistence barcode $B = \{a_i = [x_i, y_i) : i \in I\}$, the *persistent entropy* H of the filtered simplicial complex is calculated as follows:

$$H = - \sum_{i \in I} p_i \log(p_i)$$

where $p_i = \frac{\ell_i}{L}$, $\ell_i = y_i - x_i$, and $L = \sum_{i \in I} \ell_i$. Note that, when topological noise is present, for each dimension of the persistence barcode, there can be more than one interval, denoted by $[x_i, y_i)$, with $i \in I$. This is equivalent to say that, in the persistent diagram, the point $[x_i, y_i)$ could have multiplicity greater than 1 (see [10, page 152]). In the case of an interval with no death time, $[x_i, \infty)$, the corresponding barcode $[x_i, m)$ will be considered, where $m = \max\{F\} + 1$. Note that the maximum persistent entropy corresponds to the situation in which all the intervals in the barcode are of equal length. In that case, $H = \log n$ if n is the number of elements of I . Conversely, the value of the persistent entropy decreases as more intervals of different length are present. The stability theorem for persistent entropy [28] lets to compare the entropies computed from the same simplicial complex but equipped with different filtrations.

For example, let's compute the persistent entropy for the piecewise simplicial complexes in Figure 7. The persistent entropy of the space is computed as follows. The maximum filter value is 2, so the symbol " ∞ " representing the persistent topological feature is substituted with the value $m = 3$. Then, the barcode is formed by two lines with lengths $\ell_1 = 1$ and $\ell_2 = 3$, respectively. So the total length $L = 1 + 3 = 4$, for each line the probability is given by $p_1 = 1/4$ and $p_2 = 3/4$, and finally the persistent entropy is $H = 0.5623$.

3 A new topology based classifier of epilepsy

In this paragraph we describe a new methodology, that is based on Topological Data Analysis, for the analysis and classification of EEG signals. The methodology works on an EEG signal and it can be divided in three steps: in *Step I* we preprocess the input. In *Step II* we convert the input into a collection of piecewise complexes that are characterized by their

persistent homology and persistent entropy. The input EEG signal is then characterized by the average of the persistent entropy of its components. The average value is used for the classification. In *Step III* we perform two experiments. In the first experiment we aim to group brains by looking at their homological invariants, while in the second one we intend to identify the channels that characterize an EEG record. For this task we use the Vietoris-Rips complexes.

More formally, given an EEG signal S that is composed by n one-dimensional time varying signals, s.t. $S = \{s_1, s_2, \dots, s_n\}$

Step I - preprocessing

For each $s_i \in S$:

- apply a lowpass filter for reducing the noise
- decrease the sampling rate but preserving its main features. This step is required for optimizing the complexity (both temporal and spatial) of the methodology⁴. We dubbed the new decimated signal with sd_i .

Step II - classification

For each sd_i :

- obtain a piecewise complex pw_i , see Section 2.1
- describe pw_i with its persistent homology and its persistent entropy

Now the input signal S is represented by a vector of n values of persistent entropy. Compute the average of this vector. The average value of persistent entropies is a one dimensional feature that is able to differentiate signal by looking at their shapes [28].

Step III - The role of generators

In this task a brain is characterized by 23 points p_i and each p_i is within the \mathbb{R}^n space equipped with a distance. For the sake of clarity, in this experiment $p_i = \{v_{i,t_1}, v_{i,t_2}, \dots, v_{i,t_n}\}$, where v_{i,t_n} is the voltage recorded by the i -th sensor at time t_n .

- compute the pair-wise standardized euclidean distances among the $23 \times n$ sd_i signals
- from this metric space compute the Vietoris-Rips complex
- compute the persistent homology of the Vietoris-Rips complex

⁴The conversion of a signal into a piecewise complex has a temporal complexity equal to $\Theta(n)$. Also, a piecewise simplicial complex has n 0-simplices and $n - 1$ 1-simplices, where n is the number of temporal points within the input signal where n is the number of temporal points within the input signal. The complexity of the computation of the persistent homology, in JavaPlex, has in the worst case a complexity of $O(m^3)$ and in the average case $\Theta(m^2)$ Where m is the number of simplices.

- list the generators of higher dimensional holes (see Section 2.2) and study their frequency. These generators are central for the dynamics recorded by the input signal S .

4 Results and Final Remarks

In this section we report on the application of the methodology described in Section 3 to the EEG signals. We implemented the methodology in Matlab and we used the JavaPlex package for computing the persistent homology [31].

4.1 Step I - preprocessing

For the *preprocessing step* we used the MATLAB command “decimate”⁵. The lowpass filter is characterized by the normalized cutoff frequency 0.8/r and passband ripple 0.05 dB. We tested the performance of our systems with a decimation factor of 10 and 100. The results are the same in both cases. In the following we report the results obtained with the analysis with decimation factor 100, then in our setting a EEG is composed by 23 channels \times 9216 time points.

4.2 Step II - classification

We apply the methodology described in the paragraph 2.1 for converting the decimated signals into piecewise complexes, i.e. for creating the topological space, and the calculus of their persistent homology, the PE is derived. In Figure 8 we plot the comparisons between two pairs of healthy an unhealthy patients. In the first pair (top) it is not possible to observe a strong separation among the values of the normalized persistent entropy, conversely the last pair (bottom) shows good separation.

From Figure 9 we observe the class of epileptic patients is characterized by a peak of 313 elements in the range of normalized persistent entropies : $[0.942, 0.967]$ and center value 0.955, while the healthy patients are characterized by the biggest bin containing 247 elements with normalized persistent entropies values in $[0.930, 0.948]$ and center value 0.939. This first attempt does not immediately reveal any meaningful information for characterizing the two classes. Conversely, the global analysis of the EEG pinpoints out information that otherwise can not be detected. For this reason, instead of dealing with the analysis of the normalized persistent entropies for the single channel, we are focused on the study of the average value of the 23 values of persistent entropy. In Figure10 we plot for each patient the average value of its normalized persistent entropy. It is well evident that there is a strong separation between the two populations. We confirm the separation between the two classes by the statistical Wilcoxon test: $p\text{-value} = 1.8346e\text{-}36$, confidence

⁵<http://it.mathworks.com/help/signal/ref/decimate.html>

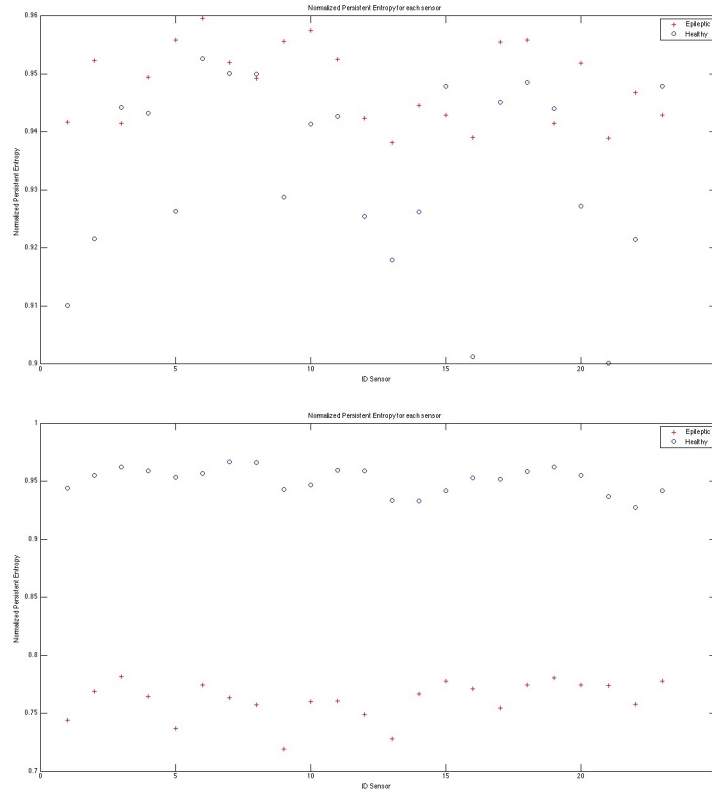


Figure 8: Normalized persistent entropy for the 23 signals within an EEG signal for epileptic patient (red) and withing a EEG signal for healthy patient (blue). Top: weak separation between the healthy and the unhealthy patients. Bottom: strong separation between the healthy and the unhealthy patients.

interval = [1.6942, 1.9675]. We used the Wilcoxon test instead of the “t-test” because the two populations are not normally distributed. For having a comparison with a well establish method already used in the analysis of time series, we also calculated the Sample Entropy [30, 24] of the data. The results can be seen in Figure 11. In this case sample entropy failed to separate the epileptic from the healthy. From the results of the PE we trained a supervised classifier. We randomly divided our dataset in two subsets: training and testing, of 70% and 30% respectively by applying a k-fold cross validation ($k = 10$).

From the study of the ROC curve (see Figure 12) we obtained that the best threshold for the separation of the two classes is: $\theta = 0.8754$. It means, that given a new EEG if its normalized persistent entropy is less than θ then it is pathological, otherwise it is healthy. The area under curve (AUC) is 97.2%

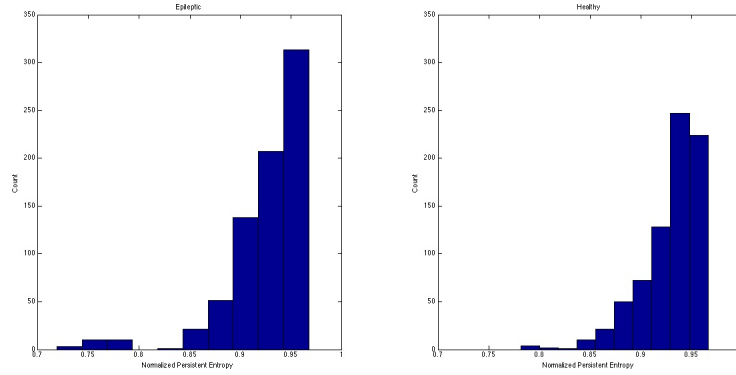


Figure 9: Histogram of the normalized persistent entropies for each sensor (23×66) for the two classes of patients (left: epileptic, right: healthy). The elements are sorted into 10 equally spaced bins along the x-axis between the minimum and maximum values of normalized persistent entropy.

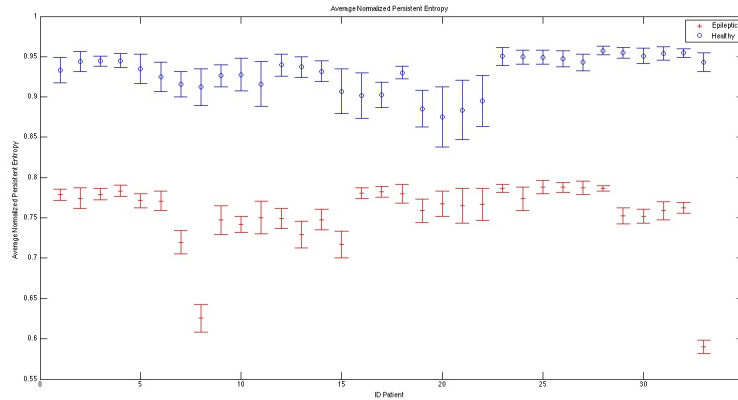


Figure 10: Average value of normalized persistent entropy over the 23 signals within the EEG for epileptic patients (red) and for healthy patients (blue). The error bars are the standard deviation.

4.3 Step III - The role of generators

The last step of the methodology aims to identify the most important sensors for describing the brain activities recorded by an EEG analysis. We accomplish this task by analyzing the Vietoris-Rips associated to each patient [15]. For each patient we compute the pair-wise standardized euclidean distance among the decimated signals (see Figure 13). Standardized

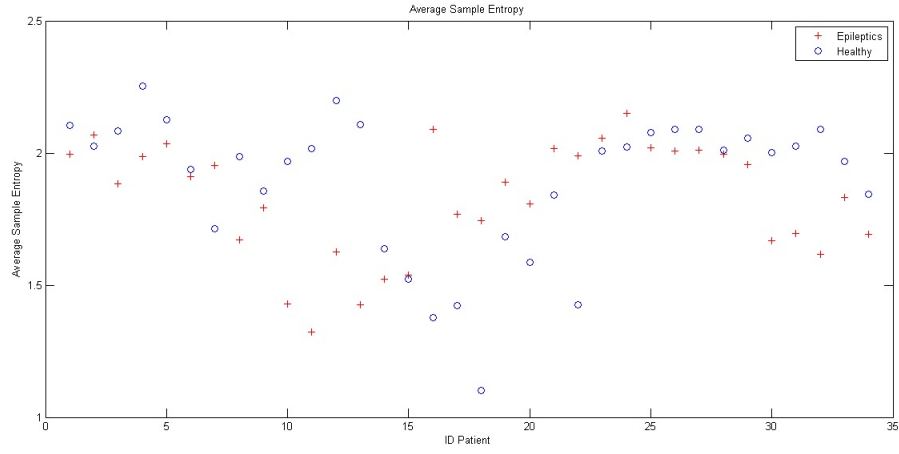


Figure 11: Average value of Sample entropy over the 23 signals within the EEG for epileptic patients (red) and for healthy patients (blu).

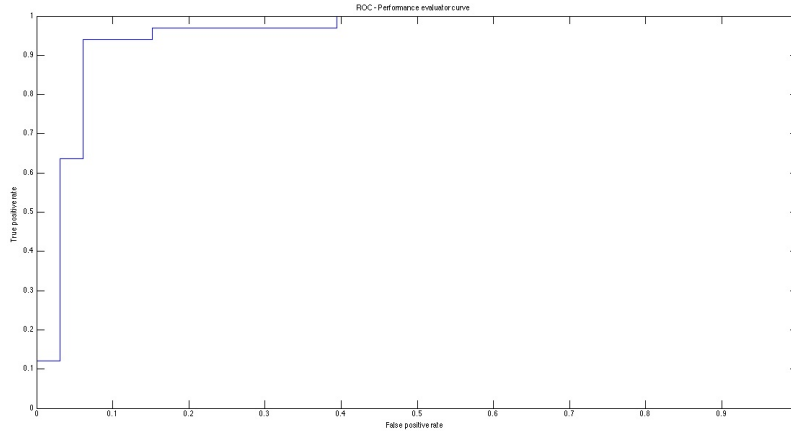


Figure 12: Receiver Operating Characteristic Curve with Area Under Curve of 97.2%.

euclidean distance is able to give better performance when data set contains heterogeneous scale variables and it is defined in Equation 1.

$$d(sd_1, sd_2) = \sqrt{\sum_{j=1}^J \left(\frac{sd_{1,j}}{s_j} - \frac{sd_{2,j}}{s_j} \right)^2} \quad (1)$$

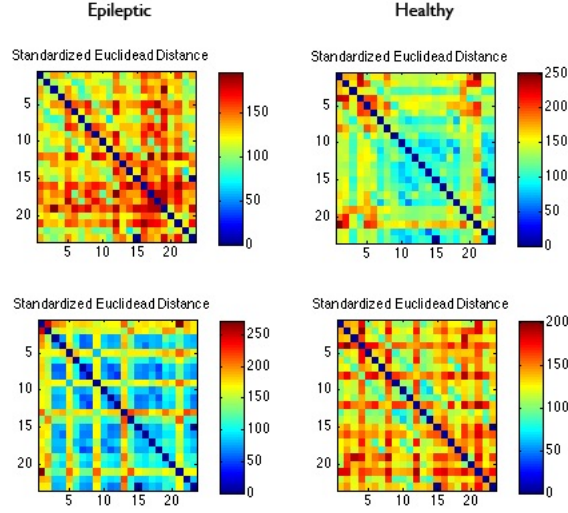


Figure 13: Four mosaics representing the pair-wise standardized euclidean distances among the 23 sensors of 4 EEG. Left: epileptic patients, right: healthy patients. From the analysis of the mosaics it is not possible to identify any patterns characterizing the two classes.

Where $sd_{1,j}$ means the j -th component of decimated signal 1, s_j is the sample standard deviation of the j -th variable. From each metric space we compute the Vietoris-Rips and we analyzed their persistent homology and the generators of the persistent higher dimensional holes. In this setting the domain of the filter function coincides with the range of the pair-wise distances and we use as upper bound of the filtration the longest distances among sensors. In Figure 14 we report two examples of persistent barcodes for one epileptic patient and for one healthy patient. We analyze all the 66 barcodes and for each patient we extract the number of β_1 (see Table 4.3) and the generators of the holes (see Figure 15). From Table 4.3 it is possible to deduce that there are less epileptic patients without β_1 than healthy patients (3 vs. 12). It means that generally the epileptic patients are characterized by β_1 . Unfortunately, this result can not be used for describing the two classes because the Wilcoxon test failed (p-value = 0.6946). However, the reader can recognize that the epileptic patients are principally characterized by 3 sensors, respectively with IDs: 1, 2 and 5 (see Figure 3 for matching the positions of the sensors). While the healthy patients are strongly characterized by sensor number 2 and also by sensors numbers: 1, 3, 7, 10 and 13. A potential interpretation is that, at least for this cohort, the epileptic events are on load of a few cortical columns, while conversely, healthy patients were stressed with stimuli that are recognized by more cortical columns. We argue this seminal result should be further investigated and if it will be validated it might be used for defining ad-hoc treatment.

As future work we aim to exploit some authors results to characterize the synchronization of n-neurons interaction in term of oscillatory models (e.g. Kuramoto model)

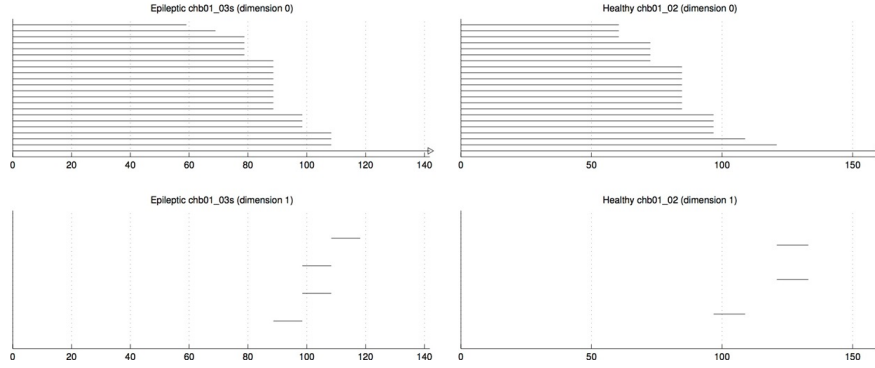


Figure 14: Example of persistent barcodes. Top: homological dimension 0 (β_0), bottom: homological dimension 1 (β_1). Left: epileptic patient, right: healthy patient.

$\#\beta_i$	Epileptic	Healthy
0	3	12
1	14	7
2	5	8
3	8	3
4	1	1
5		2

[2]. We are currently investigating the role of the simplices that generate the topological holes in the synchronization among the neurons. Our work intend to use the the extended Kuramoto model as backbone for the formal interpretation of our numerical results.

The final aim of the proposed approach in detecting the emerging behavior of epileptic seizures is to pave the ground for an automatic tool for supporting the clinicians during the diagnosis of epilepsy.

Disclosure/Conflict-of-Interest Statement

The author thanks Dr. Giovanna Viticchi and Riccardo Ricciuti from AOU - Ospedali Riuniti di Ancona (Italy), for valuable discussions on medical aspects about epilepsy. The financial support of this paper was provided by the Future and Emerging Technologies (FET) program within the Seventh Framework Programme (FP7) for Research of the European Commission, under the FET-Proactive grant agreement TOPDRIM, number FP7-ICT-318121. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

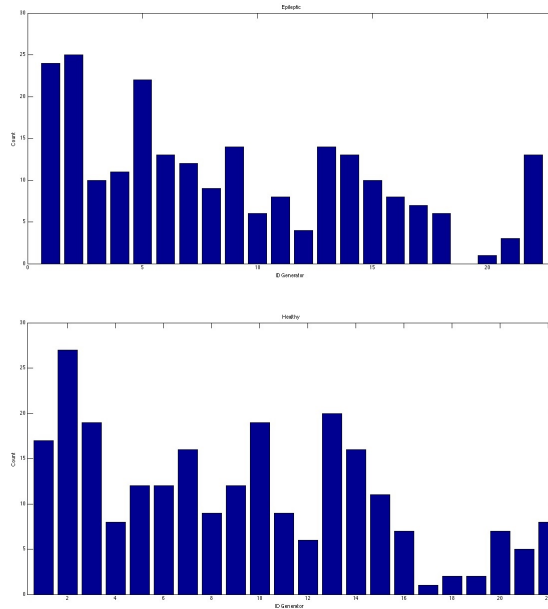


Figure 15: Bar plot of the generators belonging to higher dimensional holes for 33 epileptic patients (top) and 33 healthy patients (bottom).

References

- [1] Henry Adams and Andrew Tausz. Javaplex tutorial, 2011.
- [2] Ezio Bartocci, Flavio Corradini, Emanuela Merelli, and Luca Tesei. Model checking biological oscillators. *Electr. Notes Theor. Comput. Sci.*, 229(1):41–58, 2009.
- [3] Jacopo Binchi, Emanuela Merelli, Matteo Rucco, Giovanni Petri, and Francesco Vaccarino. jholes: A tool for understanding biological complex networks via clique weight rank persistent homology. *Electronic Notes in Theoretical Computer Science*, 306:5–18, 2014.
- [4] Gunnar Carlsson, Afra Zomorodian, Anne Collins, and Leonidas Guibas. Persistence barcodes for shapes. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 124–135. ACM, 2004.
- [5] Joseph Minhow Chan, Gunnar Carlsson, and Raul Rabadan. Topology of viral evolution. *Proceedings of the National Academy of Sciences*, 110(46):18566–18571, 2013.

- [6] Harish Chintakunta, Thanos Gentimis, Rocio Gonzalez-Diaz, Maria-Jose Jimenez, and Hamid Krim. An entropy-based persistence barcode. *Pattern Recognition*, 48(2):391–401, 2015.
- [7] Vin de Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(339-358):24, 2007.
- [8] Valentina F Domingues, Cinzia Nasuti, Marco Piangerelli, Luísa Correia-Sá, Alessandro Ghezzi, Marina Marini, Provvidenza M Abruzzo, Paola Visconti, Marcello Gius-tozzi, Gerardo Rossi, et al. Pyrethroid pesticide metabolite in urine and microelements in hair of children affected by autism spectrum disorders: A preliminary investigation. *International journal of environmental research and public health*, 13(4):388, 2016.
- [9] Herbert Edelsbrunner and John Harer. Persistent homology-a survey. *Contemporary mathematics*, 453:257–282, 2008.
- [10] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- [11] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.
- [12] L. D. Iasemidis, P. Pardalos, J. C. Sackellares, and D. S. Shiau. Quadratic Binary Programming and Dynamical System Approach to Determine the Predictability of Epileptic Seizures. *Journal of Combinatorial Optimization*, 5:9–26, 2001.
- [13] L. D. Iasemidis and J. C. Sackellares. REVIEW : Chaos Theory and Epilepsy. *The Neuroscientist*, 2(Ldi):118–126, 1996.
- [14] Abasiofiok M Ibekwe, Jincal Ma, David E Crowley, Ching-Hong Yang, Alexis M Johnson, Tanya C Petrossian, and Pek Y Lum. Topological data analysis of Escherichia coli O157: H7 and non-O157 survival in soils. *Frontiers in cellular and infection microbiology*, 4, 2014.
- [15] Svetlana Lockwood and Bala Krishnamoorthy. Topological features in cancer gene expression data. *arXiv preprint arXiv:1410.3198*, 2014.
- [16] Kaushik Majumdar, Pradeep D. Prasad, and Shailesh Verma. Synchronization implies seizure or seizure implies synchronization? *Brain Topography*, 27(1):112–122, 2014.
- [17] Patrick E McSharry, Leonard a Smith, and Lionel Tarassenko. Comparison of predictability of epileptic seizures by a linear and a nonlinear method. *IEEE transactions on bio-medical engineering*, 50(5):628–633, 2003.
- [18] Emanuela Merelli, Matteo Rucco, Marco Piangerelli, and Daniele Toller. A topological approach for multivariate time series characterization: the epilepsy case study. *Proc.*

9th EAI Conference on Bio-inspired Information and Communications Technologies (BICT 2015), 2015.

- [19] Emanuela Merelli, Matteo Rucco, Peter Sloot, and Luca Tesei. Topological characterization of complex systems: Using persistent entropy. *Entropy*, 17(10):6872–6892, 2015.
- [20] J.R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, MA, 1984.
- [21] Cinzia Nasuti, Stefano Ferraro, Rita Giovannetti, Marco Piangerelli, and Rosita Gabbianelli. Metal and microelement biomarkers of neurodegeneration in early life permethrin-treated rats. *Toxics*, 4(1):3, 2016.
- [22] Jose A Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15(3):799–838, 2015.
- [23] G Petri, P Expert, F Turkheimer, R Carhart-Harris, D Nutt, PJ Hellyer, and Francesco Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101):20140873, 2014.
- [24] Joshua S Richman and J Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6):H2039–H2049, 2000.
- [25] Osvaldo A Rosso, Susana Blanco, Juliana Yordanova, Vasil Kolev, Alejandra Figliola, Martin Schürmann, and Erol Başar. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of neuroscience methods*, 105(1):65–75, 2001.
- [26] Matteo Rucco, Filippo Castiglione, Emanuela Merelli, and Marco Pettini. Characterisation of the idiosyncratic immune network through persistent entropy. In *Proc. of 11th European Conference on Complex Systems (ECCS 2014)*, Proceedings in Complexity, pages 117–128. Springer, 2015.
- [27] Matteo Rucco, Enrico Concettoni, Cristina Cristalli, Andrea Ferrante, and Emanuela Merelli. Topological classification of small dc motors. In *Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), 2015 IEEE 1st International Forum on*, pages 192–197. IEEE, 2015.
- [28] Matteo Rucco, Rocio Gonzalez-Diaz, Maria-Jose Jimenez, Nieves Atienza, Enrico Concettoni, Cristina Cristalli, Andrea Ferrante, and Emanuela Merelli. A new topological entropy-based approach for measuring similarities among piecewise linear functions. *In print, Signal Processing, Elsevier*, 2016.

- [29] Sabato Santaniello, Samuel P Burns, Alexandra J Golby, Jedediah M Singer, William S Anderson, and Sridevi V Sarma. Quickest detection of drug-resistant seizures: an optimal control approach. *Epilepsy & behavior : E&B*, 2011.
- [30] Yuedong Song, Pietro Liò, et al. A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine. *Journal of Biomedical Science and Engineering*, 3(06):556, 2010.
- [31] Andrew Tausz, Mikael Vejdemo-Johansson, and Henry Adams. javaplex: a research platform for persistent homology. *Book of Abstracts Minisymposium on Publicly Available Geometric/Topological Software*, 7, 2012.
- [32] Dane Taylor, Florian Klimm, Heather A Harrington, Miroslav Kramár, Konstantin Mischaikow, Mason A Porter, and Peter J Mucha. Topological data analysis of contagion maps for examining spreading processes on networks. *Nature communications*, 6, 2015.
- [33] Qawi K Telesford, Sean L Simpson, Jonathan H Burdette, Satoru Hayasaka, and Paul J Laurienti. The Brain as a Complex System: Using Network Science as a Tool for Understanding the Brain. *Brain Connectivity*, 1(4):295–308, October 2011.
- [34] Adrian P. Sheppard Vanessa Robins, Peter John Wood. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Trans. Pattern Anal. Mach.*, 33(8):1646–1658, 2011.
- [35] Yue Yang and Huijie Yang. Complex network-based time series analysis. *Physica A: Statistical Mechanics and its Applications*, 387(5-6):1381–1386, February 2008.
- [36] Afra Zomorodian. Topological data analysis. In *Advances in Applied and Computational Topology. Proceedings of Symposia in Applied Mathematics*, volume 70, pages 1–39, 2007.
- [37] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.